Increased Math Achievement in Elementary Students Participating in
JUMP Math's 2015-16 National Book Fund Program.
Beverley Murray, Ph.D.
January 11, 2017

Executive Summary

JUMP Math characterizes its approach to math instruction as *guided discovery*, a combination of direct instruction, discovery learning, and varied practice.[1] The program breaks down math problems into incremental steps and advocates mastery of simpler concepts before advancement to more complex concepts. Scaffolding of math problems is widely used to assist with independent learning. The program also promotes the importance of building student confidence and the notion that all students are capable of learning mathematics with appropriate supports.[2] Components of the program include professional development; *Teacher Resources* composed of lesson plans, quizzes/tests, and answer keys; *SMART Lesson Materials* for use with interactive white boards; and student *Assessment & Practice* books.

To evaluate the growth of students using the JUMP Math program, math achievement was assessed in both the fall and spring for students in grades 4 to 7 who participated in JUMP Math's 2015-16 National Book Fund (NBF) program. A total of 208 students in ten classrooms completed the math computation subtest of the *Wide Range Achievement Test, Fourth Edition* (WRAT-4) in October 2015 and May 2016. Average student growth in math achievement was 2.4 times that of the WRAT-4 standardization sample, and the mean standard score in the spring (M = 97.0, SD = 13.3) was significantly higher than the mean standard score in the fall (M = 92.4, SD = 13.8), paired t(207) = 6.7, p < 0.001. The corresponding percentile rank of students increased from the 30th percentile in the fall to the 42nd percentile in the spring. The number of students scoring 'above average' or higher increased by 64% in the spring (41 students) compared to the fall (25 students). The number of students scoring 'below average' decreased by 18% in the spring (76 students) compared to the fall (93 students). We cannot know for certain whether the increased growth in math achievement relative to the WRAT-4 was due solely to the JUMP Math program because this study did not employ randomized control and treatment groups. By using a standardized test with alternate forms, however, we reduced the potential impact of several confounding variables making it likely that the JUMP Math program played a significant role.

---

[1] http://www.jumpmath.org/
[2] Mighton, J. (2004). *The myth of ability: nurturing mathematical talent in every child.* Toronto: House of Anansi Press.

<u>Background</u>

Every year, JUMP Math's National Book Fund Program awards free JUMP Math resources to classrooms across Canada. The lead funder of this program is the TD Bank Group. To be considered for the award, school principals and teachers must submit an application in which they describe their community and the needs of their students. Priority for awards is given to schools serving high-need communities where student achievement in mathematics is below national standards. In the 2015-16 school year, JUMP Math's National Book Fund Program awarded resources to over 4,500 students in 161 classrooms across 9 Canadian provinces/territories (AB, BC, MB, NB, NL, NS, NU, ON, and QC).

To assess the growth in math achievement for students participating in the NBF program, non-blended, grade 4 to 7 classrooms were selected for testing. Teachers were asked to administer the math computation subtest of the *Wide Range Achievement Test, Fourth Edition* (WRAT-4)[3] to their students in October 2015 and again in May 2016. Teachers were sent two alternate forms of the WRAT-4, designated the green form and the blue form, consisting of different questions but considered equally difficult. Detailed instructions on how to administer the test and return envelopes were provided to each teacher. In the fall and again in the spring, teachers were sent test forms pre-labelled with their students' names; half of students received the blue form and half received the green form. For the spring testing, students received the alternate coloured form. Completed tests were sent back to JUMP Math and scored by a qualified teacher and the researcher. Standard scores were determined for each student in the spring and fall by looking up their raw test score in conversion tables corresponding to the student's grade, test form (blue versus green), and time of testing (fall versus spring).

---

[3] Wilkinson G. & Robertson G. (2006). *Wide range achievement test (4th ed.).* Lutz, FL: Psychological Assessment Resources, Inc.

Results

Teachers from 10 classrooms administered the WRAT-4 in both the fall and spring of the 2015-16 school year. Standard scores were determined for the 208 students who completed the tests in both the fall and spring. The mean standard score in the spring (M^4 = 97.0, SD^5 = 13.3) was significantly higher than the mean standard score in the fall (M = 92.4 SD = 13.8), paired t (207) = 6.7, p < 0.001 (see Figure 1). We would expect the students in the 2015-16 NBF to have the same standard score in the fall and spring if their math achievement had increased at the same rate as the WRAT-4 standardization sample. The fact that their mean standard score was significantly higher in the spring indicates that their math achievement grew at a rate higher than the WRAT-4 standardization sample. The corresponding percentile rank of students (relative to the WRAT-4 standardization sample) increased from the 30th percentile in the fall to the 42nd percentile in the spring. Using the published standard deviation for the WRAT-4 (SD = 15), this increase in standard score corresponds to an effect size of 0.31 ((97.0 – 92.4)/15).
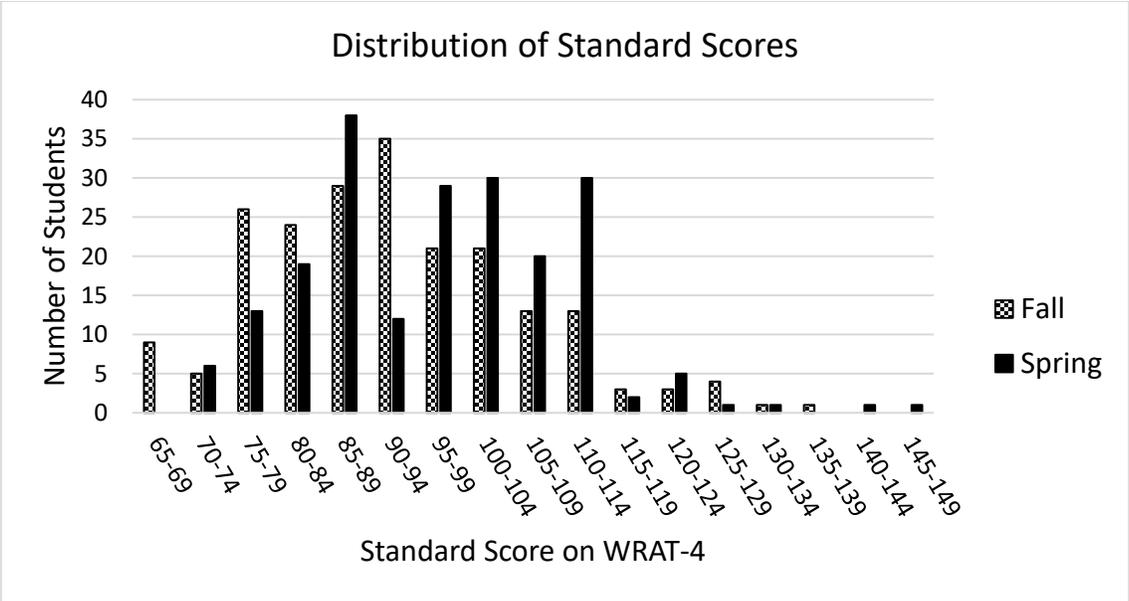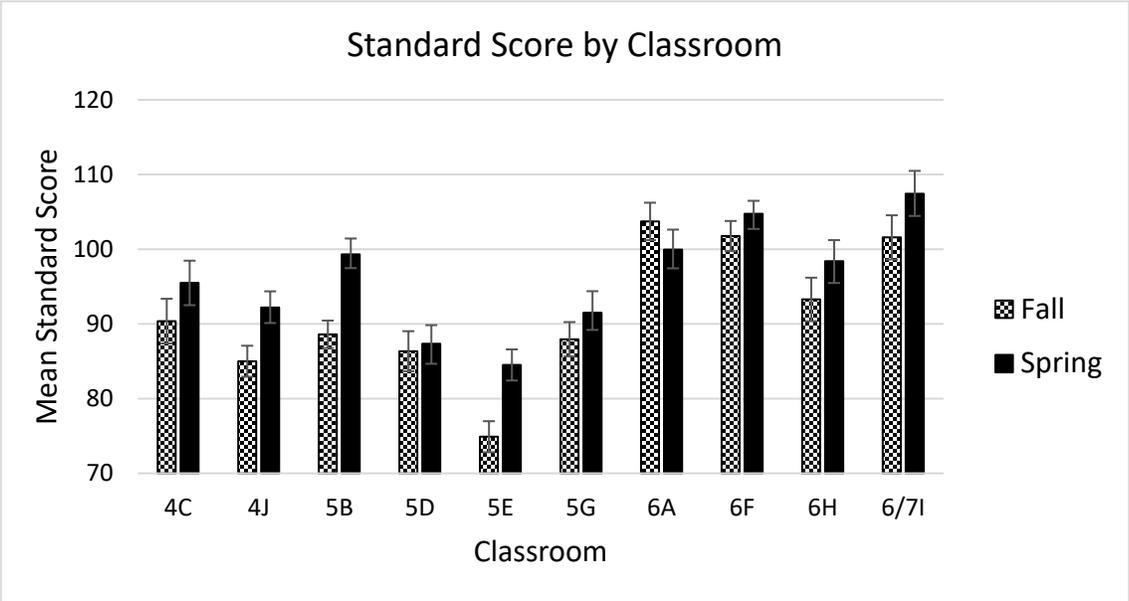
Figure 1:
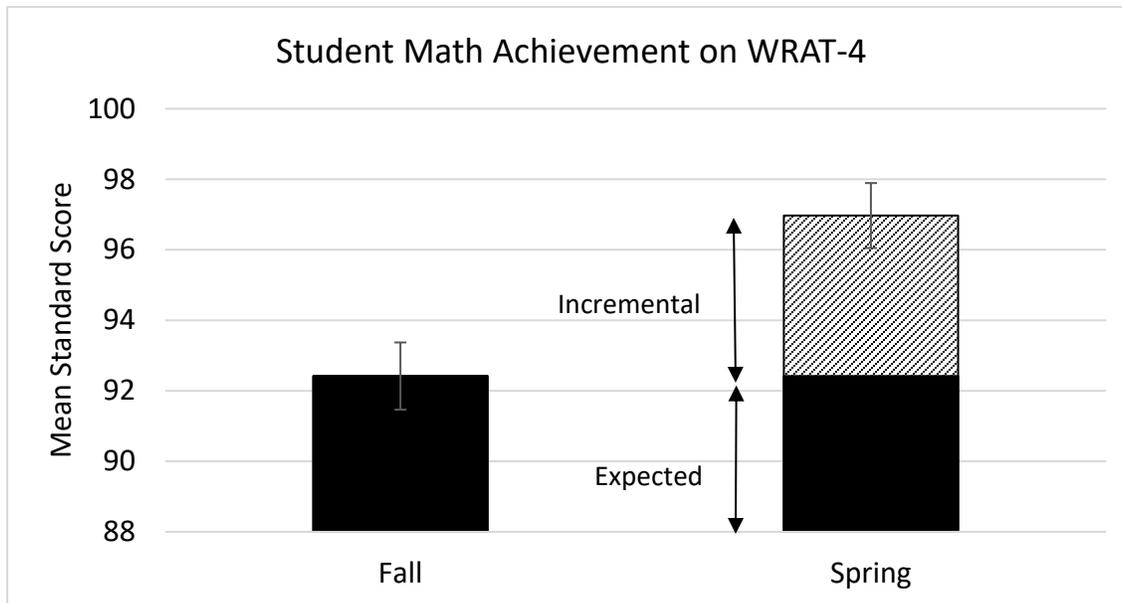
The frequency distributions of standard scores obtained in the fall and spring are shown below in Figure 2. The distributions include only those students (N=208) who completed either a blue or green test in the fall and then completed the alternate coloured test in the spring (students who completed the same test in both the fall and spring were excluded from the analysis). The graph illustrates that the distribution of scores obtained in the spring (solid bars) is shifted to the right (towards higher scores) compared to the distribution obtained in the fall (hatched bars). The median score increased from 90 in the fall to 96 in the spring. The number of students scoring 'above average' or higher (110 or higher) increased by 64% in the spring (41 students) compared to the fall (25 students). The number of students scoring 'below average' (89 or lower) decreased by 18% in the spring (76 students) compared to the fall (93 students).

Figure 2:

Mean standard scores in the fall and spring for each of the ten classrooms are shown in Figure 3 (error bars for all graphs denote the standard error of the mean (SEM)). One of the classrooms (6/7I) was a blended grade (6/7) classroom whereas the remaining classrooms were single grade classrooms. Mean standard scores in the fall ranged from 74.9 (classroom 5E) to 103.7 (classroom 6A) whereas mean standard scores in the spring ranged from 84.5 (classroom 5E) to 107.4 (classroom 6/7I). Mean standard scores increased in 9 out of 10 classrooms, with gains ranging from 1.2% (classroom 7J) to 12.8% (classroom 5E); one classroom had a 3.6% decrease in mean standard score (classroom 6A).

Figure 3:

Mean standard scores in the fall and spring for each grade are shown in Figure 4. Mean standard scores in the fall ranged from 85.8 (grade 5) to 104 (grade 7) whereas mean standard scores in the spring ranged from 91.5 (grade 5) to 107.9 (grade 7). Percent increases in mean standard score ranged from 2.3% (grade 6) to 7.3% (grade 4).
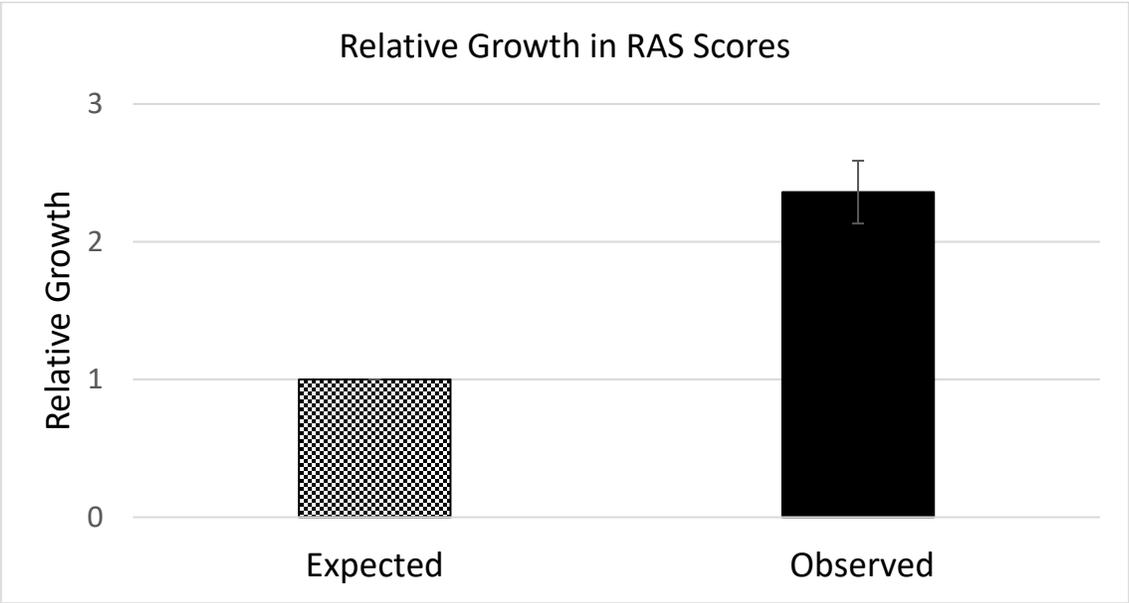
Figure 4:
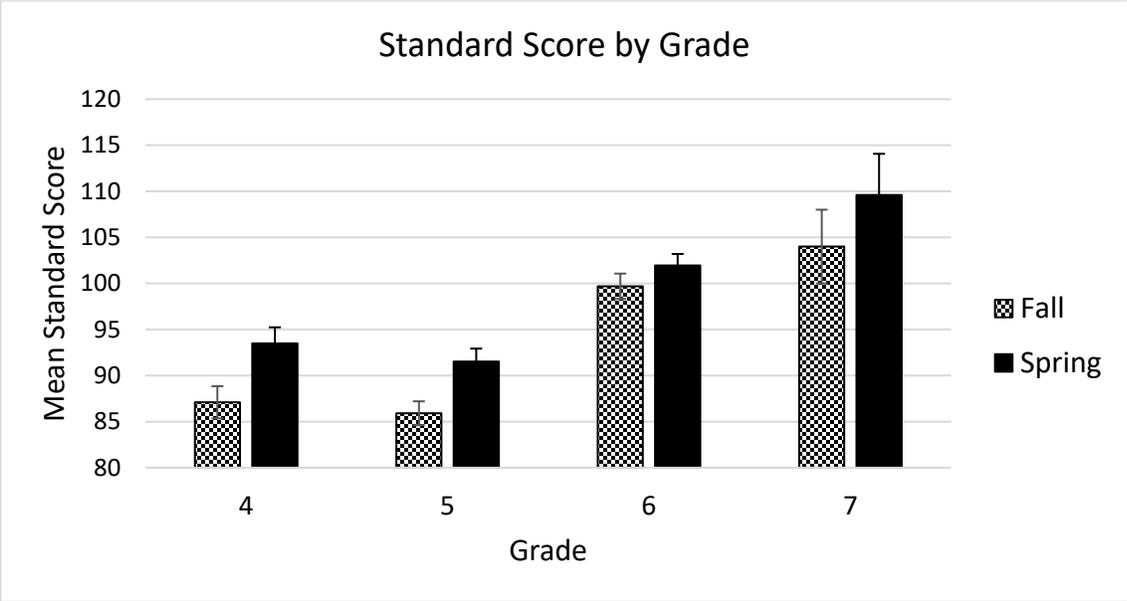
Each student's raw score on the WRAT-4 math test was also converted to a Rasch Ability Scaled (RAS) score using conversion tables for the blue and green forms of the WRAT-4. A student's RAS score will increase over time as their math achievement (raw score) increases; the RAS score is therefore well suited to measuring growth in student achievement from one time to another. In contrast, standard scores will remain constant over time if the student grows at the same rate as the standardization sample.

We defined *observed growth* as the difference between a student's RAS score in the spring and their RAS score in the fall (observed growth = RAS score in spring – RAS score in fall). *Expected growth* was also calculated for each student by subtracting their RAS score in the fall from their expected RAS score in the spring (expected growth = expected RAS score in spring – RAS score in fall). The expected RAS score in the spring was determined for each student by calculating the raw score in the spring that would result in the same standard score the student had obtained in the fall. We defined each student's *relative growth* in math achievement (relative to the WRAT-4 standardization sample) as the ratio of observed growth to expected growth (relative growth = observed growth/expected growth). Thus, a student with a relative growth score of 1 grew at the same rate as students from the WRAT-4 standardization sample with the same fall standard score. On average, math achievement of students in the 2015-16 NBF grew at 2.4 times the rate of the WRAT-4 standardization sample (Figure 5).

Figure 5:

A scatter plot of each student's standard score in the fall versus their observed growth on the WRAT-4 is shown in Figure 6. The regression line through these points has a negative slope (-0.09) and a correlation coefficient (R^2) of 0.15. This negative correlation is consistent with the phenomenon of regression-to-the-mean (RTM) and is unlikely to reflect any selective effect of the JUMP program on low-achieving students. The implications of RTM are discussed below (see page 11).

Figure 6:

Regression to the Mean

$y = -0.0931x + 10.97$
$R^2 = 0.1515$

Observed Growth

Standard Score in Fall
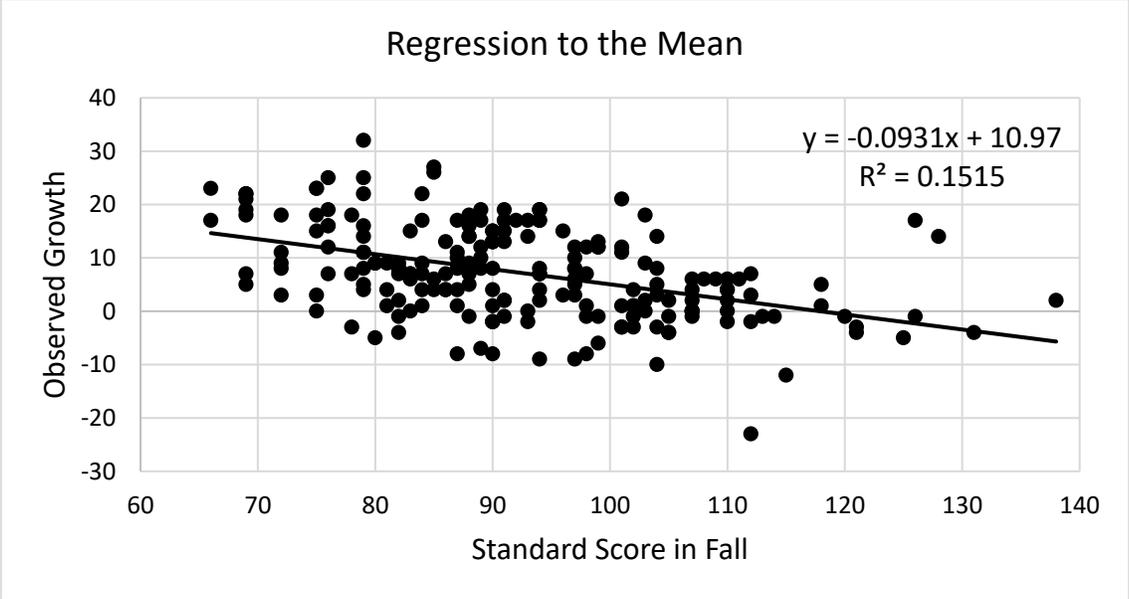
Discussion

Standard scores are a useful measure for comparing student achievement on a standardized test; a student with a standard score of 100 has achieved a score equal to the mean score of the sample of students used to standardize the test. The corresponding percentile rank is 50%; half of the students in the standardization sample scored above 100 and half scored below 100. Students in the WRAT-4 standardization sample were tested in both the fall and spring of the school year. Thus, a student who demonstrates the same growth rate as the standardization sample will achieve the same standard score in the fall and spring of the school year. Students participating in JUMP Math's 2015-16 National Book Fund Program showed significant increases in mean standard score in the spring (M = 97.0, SD = 13.3) compared to the fall (M = 92.4 SD = 13.8). The corresponding percentile rank of NBF students increased from the 30th percentile in the fall to the 42nd percentile in the spring. The number of students scoring 'above average' or higher (110 or higher) increased by 64% in the spring (41 students) compared to the fall (25 students). The number of students scoring 'below average' (89 or lower) decreased by 18% in the spring (76 students) compared to the fall (93 students). On average, students participating in the 2015-16 National Book Fund grew at 2.4 times the rate of the WRAT-4 standardization sample.

Whereas the distribution of standard scores for the WRAT-4 standardization sample has a normal distribution (i.e. a symmetric, bell-shaped curve) centred on a standard score of 100, the distribution of standard scores obtained in the present study is positively skewed, particularly in the fall, and centred on a lower standard score (median standard score in the fall was 90). A positive skew occurs when the right tail of the distribution is longer than the left tail. Our skewed distribution could be due to the selection process for the National Book Fund Program. Classrooms that were selected for the program were (mostly) from high-need communities where math achievement was below national standards. In the spring the distribution of scores shifted towards higher standard scores and was less skewed.

JUMP Math has used the WRAT-4 to assess student math achievement since the 2011-12 school year[6]. A comparison of the test results for the past 5 years is shown in Table II. We observed a step increase in student growth between the 2011-12 and 2012-13 NBF (1.8 vs 2.8, respectively) that remained high in 2013-14 (2.5), 2014-15 (2.9), and 2015-16 (2.4). The step increase could be due to policy changes that were implemented in 2012-13, the most significant of which was requiring all participating teachers to complete a JUMP Math professional development session prior to the start of the school year. In contrast, less than half of the teachers selected for classroom testing in the 2011-12 NBF had completed JUMP Math professional development by mid-October 2011 (only 8 of

[6] Murray, B. (2013, September 18). *Increased math achievement in grade 3 and 6 students participating in JUMP Math's 2011-12 National Book Fund Program.* Retrieved from http://www.jumpmath.org/cms/sites/default/files/Student%20Achievement%20From%202011-12%20JUMP%20Math%20Book%20Fund%20%282013%29.pdf

18 teachers). In addition, changes were made in the process for assessing NBF applications which may have improved our ability to identify high-needs classrooms. This may account for the increase in the percentage of low-scoring students participating in the 2012-13 NBF. Either of these policy changes could have impacted growth in student math achievement.

Table II: NBF Student Test Results

School Year	2015-16	2014-15	2013-14	2012-13	2011-12
# of students tested in both fall & spring	208	248	241	286	326
Grades tested	4 to 7	3 to 9	4	4 to 7	3 and 6
SS fall vs SS spring	92.4 vs 97.0	86.9 vs 92.3	89.6 vs 95.3	90.8 vs 94.6	96.8 vs 100.9
Percentile rank fall vs spring	30th vs 42nd	19th vs 30th	25th vs 37th	27th vs 37th	42nd vs 53rd
Average student growth relative to WRAT-4	2.4	2.9	2.5	2.8	1.8
Increase in % of students scoring 'above average' (fall vs spring)	12% vs 20%	3% vs 9%	5% vs 10%	7% vs 12%	10% vs 22%
Decrease in % of students scoring 'below average' (fall vs spring)	45% vs 37%	63% vs 45%	54% vs 35%	47% vs 36%	26% vs 20%

This study is an example of a single-group, pre- and post-test research design. This design is also referred to as "pre-experimental" because subjects have not been randomly assigned to treatment and control groups as in a true experimental design. The lack of randomized control and treatment groups in this study limits our ability to make causal inferences due to possible confounding factors. There are four well-recognized confounding factors unique to pre-experimental research studies: history, maturation, test effects, and regression-to-the-mean[7]. We have reviewed the potential impact of each of these factors and conclude that it is unlikely they can account for the statistically significant gains in math achievement obtained for the 5 years of the NBF in which student testing has been completed and analyzed. In particular, our use of a standardized test with two alternate

---

[7] Marsden, E. & Carole J. Torgerson, C.J. (2012). Single group, pre- and post-test research designs: Some methodological concerns. *Oxford Review of Education, 38*(5), 583-616.

forms eliminates any possible practice effect that may occur when students complete the same test in the fall and spring of the same school year.

One potential confounding factor that requires careful consideration in studies employing a pre-experimental design is regression-to-the-mean (RTM). RTM is a statistical phenomenon whereby a distribution of measurements (e.g. test scores) narrows with repeated observations[8]. The effect is due to the greater measurement error in the tails of the distribution. Students with very low test scores will be more likely to have higher scores on a subsequent test. Similarly, students with very high test scores will be more likely to have lower scores on a subsequent test. The effects of RTM can lead education researchers to erroneously conclude that their treatment had a greater effect for low-achieving students. To determine whether RTM was evident in this data set, the observed growth for each student (RAS score in spring – RAS score in fall) was plotted against their standard score in the fall (Figure 6). If RTM was present, students with a low standard score in the fall would tend to have a higher score in the spring (and therefore greater growth) whereas students with a high standard score in the fall would tend to have a lower score in the spring (and therefore lower growth). We would therefore expect growth to be negatively correlated with standard score in the fall (i.e. a regression line through the points would have a negative slope). The scatter plot and regression line in Figure 6 indicate a negative correlation for these data (slope = -0.09). The correlation is statistically significant, accounting for 15% of the variance in student growth ($R^2 = 0.15$).

If RTM was entirely responsible for the increases in standard score, we would expect that the shifts in standard score would be observed in the tails of the distribution, i.e. for students with the lowest and highest test scores. The impact of RTM can be summarized as an overestimation of gains for low-scoring students and an underestimation of gains for high-scoring students. Given that we observed an overall change in the mean standard score, we can conclude that other factors must have contributed to the increases in standard score and that any potential effect of RTM was most likely greatest at the tails of the distribution.

---

[8]Barnett, A.G., van der Pols, J.C. & Dobson, A.J. (2005). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology, 34*, 215–220.